Servière Adrien
Jin Alex

ENSAE 3^{rd} year
2022-2023

Project.

Bayesian Statistics.

December 2022.

Contents

1 A review of the Article

1.1 Introduction on Stochastic volatility models and motivation

When it comes to modeling financial time series, researchers develop models to take stylized facts into account. One of them is that conditional volatility is random. Indeed if y_t is the log-return of an asset, $(F_t)_t$ a filtration representing information up to t, then $E[y_t|F_{t-1}]$ is the best approximation in the least square sense with information up to time t. Therefore the quantity $y_t - E[y_t|F_{t-1}]$ is the error. In practice, the average of this squared error conditional to information up to t is $E[(y_t - E[y_t])^2|F_{t-1}]]$ (which is the conditional volatility), is stochastic. This phenomenon cannot be reproduced via usual models like ARMA, because their conditional volatility is constant. One needs to find models that make it possible to have a stochastic conditional volatility. A famous class of volatility models is the stochastic volatility models (SVM), which takes the following form in the article (we consider a discrete time version)

$$\begin{aligned} y_t &= \sqrt{h_t}\epsilon_t \\ \ln h_t &= \alpha + \delta \ln h_{t-1} + \sigma_\nu \nu_t \end{aligned}$$

with

$$(\epsilon_t, \nu_t) \sim \mathcal{N}(0, I_2)$$

Indeed with the log-volatility following an AR(1), the volatility is stochastic. The δ cefficient makes it possible to model *volatility clustering*, meaning that if the volatility was high yesterday, it is likely to be high today and vice-vera. This effect is modeled if δ is sufficiently positive (it will be the case in what follows).

However this first version suffers from several flaws that the article under study will try to counter. First given ϵ is Gaussian, y_t has thin tails, which is known to be false for log-returns. That is why the article propose to extend it to *fatter tails*. Another stylized fact that is not taken into account by the above expression is the so-called *leverage effect* in which negative (positive) shocks to the mean are likely to be associated with increases (decreases) in volatility. Mathematically, it means there is a correlation between ϵ and ν in the model. This is motivated in the article by the overwhelming evidence of a leverage effect for weekly and daily equity indices. The evidence in favor of fat-tails is very strong for daily exchange rate and equity indices, but less for weekly data.

The author of article brings the solution by first introducing a Bayesian inference Gibbs sampler algorithm for the usual SVM. Then an extension to fatter tails is provided. After that a SVM with correlated ϵ and ν is introduced before the final one that combines the two. In each of these parts, the first sub-section is about the model itself, the second is about the prior, and the third describes the algorithm.

1.2 Bayesian inference of the usual SVM

First, we implemented the MCMC algorithm JPR developed for the basic SVM model, as presented by the paper. We use MCMC methods because the law of $log(h_t)$ is intractable, thus we used a Gibbs sampler. The paper describes an a more complex version of it, that includes a step of Metropolis-Hasting combined with an accept-reject step to sample for this law.

The parameters are updated with the following posteriors for $\omega = \alpha, \delta, \sigma_\nu$:

$$\begin{aligned}
&\sigma_\nu^2 \sim Inv\Gamma(\frac{T}{2} + c_0, \frac{1}{2}\sum_{1\le t\le T}(\sigma_t - (\beta_0 + \beta_1\sigma_{t-1}))^2 + d_0) \\
&\alpha \sim \mathcal{N}(\mu, \sigma) \\
&\delta \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)\mathbf{1}_{-1\le\beta_1\le 1}
\end{aligned}$$

where,

$$\begin{aligned}
\mu &= \frac{\frac{1}{\tau^2}}{\frac{T}{\tau^2} + \frac{1}{\gamma_0}}\sum_{1\le t\le T}(\sigma_t - \beta_1\sigma_{t-1}) + \frac{\frac{1}{\gamma_0}}{\frac{T}{\tau^2} + \frac{1}{\gamma_0}}\alpha_0 \\
\sigma^2 &= \frac{1}{\frac{T}{\tau^2} + \frac{1}{\gamma_0}} \\
\tilde{\mu} &= \frac{1}{\frac{1}{\tau^2}\sum_{1\le t\le T}\sigma_{t-1}^2 + \frac{1}{\gamma_1}}\left(\frac{1}{\tau^2}\sum_{1\le t\le T}(\sigma_t - \beta_0)\sigma_{t-1} + \frac{1}{\gamma_1}\alpha_1\right) \\
\tilde{\sigma}^2 &= \left(\frac{1}{\tau^2}\sum_{1\le t\le T}\sigma_{t-1}^2 + \frac{1}{\gamma_1}\right)^{-1}
\end{aligned}$$

and the posterior of H_t is,

$$p(h_t|h_{t-1}, h_{t+1}, \alpha, \delta, \sigma_\nu, y) \propto h^{-1,5}\exp(-\frac{y_t^2 + (\log(h_t - \mu_t))^2}{2(h_t + \sigma^2}))$$

We use an inverse gamma blanket $q \sim Inv\Gamma(\phi_t, \theta_t)$ (with the parameters as described in the paper) to sample from the distribution of h_t in the Metropolis step.

1.3 Extension to fatter tails

Empirical evidences suggest that the errors are not well described by the normal distribution. Now, the paper focuses on using a fatter-tailed distibution : $\epsilon_t := \sqrt{\lambda_t} z_t$ is a $Student - t(\nu)$. This allows for flexibility through the additional term λ_t to model outliers (by using a larger value of λ_t).

We adapt the previous model to take this into account: we use a conjugate inverse gamma for $\lambda_t|\nu$.

$$\lambda_t|\nu \sim Inv\Gamma(\nu/2, 2/\nu)$$
$$\nu \sim U[3, 40]$$

We thus have the posteriors :

$$\lambda_t|y_t, h_t, \nu \sim Inv\Gamma(\frac{\nu+1}{/2}, \frac{2}{(y_t^2/h_t)+\nu}) \tag{1}$$
$$p(\nu|h, \omega,) = p(\nu) \prod_{t=1}^{T} \nu^{\nu/2} \frac{\Gamma(\nu+1/2)}{\Gamma(\nu/2)\Gamma(1/2)} (\nu + y_t^2/h_t)^{-\frac{\nu+1}{2}} \tag{2}$$

and the same for the other parameters as in the basic model by using the fact that $y_t^* := y_t \lambda_t^{-0.5} = h_t^{0.5} z_t$ with $z_t \sim \mathcal{N}(0, 1)$.

1.4 Extension to correlated noises

We are now interested in the relationship between ϵ_t and $v_t =: u_t/\sigma_\nu$. We consider the following covariance matrix : $\Sigma^* = \begin{pmatrix} 1 & \rho\sigma_\nu \\ \rho\sigma_\nu & \sigma_\nu^2 \end{pmatrix} =: \begin{pmatrix} 1 & \psi \\ \psi & \Omega + \psi^2 \end{pmatrix}$ Notice we have reparametrized the model with (Ω, ψ) and one can go back to $(\rho, \sigma_n u)$ with a simple transformation. We have the priors :

$$\Omega \sim Inv\Gamma(\nu_0, \nu_0 t_0^2)$$
$$\psi|\Omega \sim \mathcal{N}(\psi_0, \Omega/p_0)$$

and the posteriors (with the exact expression of the parameters defined in the article):

$$\Omega|\alpha, \delta, h, y \sim Inv\Gamma(\nu_0 + T - 1, \nu_0 t_0^2 + a_{22.1}) \tag{3}$$
$$\psi|\Omega, \alpha, \delta, h, y \sim \mathcal{N}(\tilde{\psi}, \Omega/(p_0 + a_1 1)) \tag{4}$$

The posterior of h_t changes and becomes :

$$p(h_t) \propto \frac{1}{h_t^{1.5+\frac{\delta\psi y_t+1}{\Omega\sqrt{h_{t+1}}}}} \exp(\frac{-y_t^2}{2h_t}(1+\frac{\psi^2}{\Omega}) - \frac{(\log h_t - \mu_t)^2}{2\Omega/(1+\delta^2)} + \frac{u_t\psi y_t}{\Omega\sqrt{h_t}}) \tag{5}$$

We will use an $Inv\Gamma(\phi_t, \theta^*)$ blanket to sample from this distribution, again the exact method is described in the paper and it would be far too long to explain it here.

1.5 Algorithm for fat-tails and correlation

We can simply combine the two models : we use the correlated model on y^*, as defined in the fat-tail section and do minor adjustment to the algorithm by sampling from (1) and (2).

2 The results

We will test these algorithms first on simulated data and then on real data from yahoo-finance: the S&P500 index of 1993 (the first availabale year). We will do $N = 3500$ steps in our Gibbs Sampler.

2.1 Simulated data

We simulated under the basic SVM with the following parameters: $\sigma_\nu = 0.4, \alpha = -0.5, \delta = 0.9$ with $T = 100$ data points and a burnin of 2000 to ensure stationarity of the AR(1) and convergence of the Gibbs sampler.

We tested all 4 algorithms on this data which can be found in the notebook provided along this report.

In the article, the authors used an Accept-reject Metropolis-Hasting within a Gibbs, we have tested (see notebook) with a basic Metropolis-Hasting to verify if it was effectively better. In fact, the pure Metropolis-Hasting algorithm struggles to find suitable values for $\ln(h_t)$. Indeed, like the article, we find an accpet-reject rate of around 20%.

For the fat-tails version, we expect to find a high ν : if the degree of freedom $\nu = +\infty$, we go back to a gaussian distribution (aka the basic model).

For the correlated model, we expect to find $\psi = 0$ and $\Omega = \sigma_\nu^2$, which would mean uncorrelated mean and variance errors (aka the basic model).

Here, we will only discuss the basic model and the full model (fat-tail + correlated) but the trajectories and posterior distribution, parameter estimation details can be found in the notebook.

2.1.1 Basic model

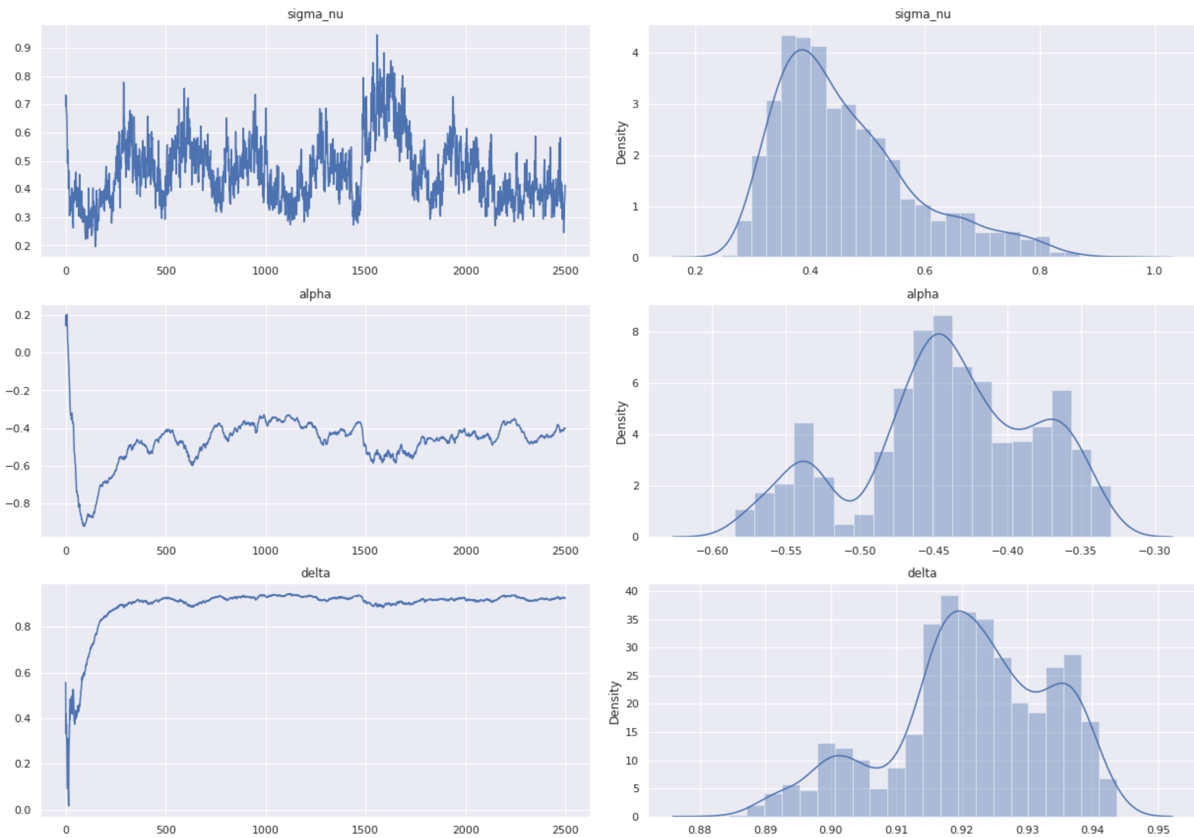

Figure 1: Trajectories and posteriors for the basic model and simulated data

Here, we can see on the trajectories that δ and α have converged towards their value, while σ_ν remains quite erratic, but still around the correct value. When we look into their distribution, we can see that the histrogram for σ_ν looks like an inverse gamma whereas for the other two, their histogram isn't quite the distribution of a truncated normal and normal distribution respectively. Perhaps we should have add iterations in our model, given the very high number of parameters in the gibbs sampler (we have to sample h_t for every t at each iteration)

We can do parameter estimation with this model and we obtain :

- σ_ν has mean 0.46 and std 0.12 (real value is 0.4)
- α has mean -0.44 and std 0.06 (real value is -0.5)
- δ has mean 0.92 and std 0.012(real value is 0.9)

These estimated values for each parameter remain close to their true value.

2.1.2 Full model

For the fat-tailed version, we obtain as expected a big value of $\nu = 40$ very early on (which is the maximum allowed by our model) and the same problems for $_\nu$ as the previous model is observed. Furthermore, λ_t remains small (mean = 1.64, std = 0.25) as expected since there shouldn't be many outliers in the simulated data.

For the correlated model, α takes a lot of time to converge and stabilise around the wrong value. δ has a very erratic behaviour in the beggining but becomes stationnary after a while. Ω is initialised with a large value and becomes stationnary very fast. As expected, ψ stays around 0. The problems in the posterior shown in the basic model remains here.

The full model combines the previous two, and also combines their problems. The trajectories and posteriors are about the same, and the same remarks can be made. In particular, we can see an order of magnitude between the estimated Ω and the real one. This result is unexpected but we didn't see any mistake in our code, but perhaps there might be one.

- Ω has mean 0.01 and std 0.0012 (real value is $0.4^2 = 0.16$)
- α has mean -0.58 and std 0.022 (real value is -0.5)
- δ has mean 0.92 and std 0.01(real value is 0.9)
- ψ has mean 0.01 and std 0.011 (real value is 0)
- ν is 40 (max allowed by model, real value is $infty$)

As we can see on the trajectory of the parameters, overall, they converge (quite slowly) towards their real value. We have used way less iteration and data due to computational power limitation ; the article used $T = 3000$ data points and $N = 30000$ iterations. One may argue that we would have needed more iterations to see the full convergence.

We have encountered some difficulties when implementing the algorithm from the paper : Some a posteriori distribution where not provided and we had to compute them ourselves ; and some of the methods where unclear : the Accept-Reject Metropolis-Hasting algorithm is not described and we had to look it up in another paper 'Accept-reject Metropolis-Hasting sampling and marginal likelihood estimation' from Siddhartha Chib. Also the way we determine s in θ^* for the correlated model (linearisation of a term around the mode) is unclear. Furthermore, apparently, the authors of the paper used the convention where their scale parameter was the inverse of the usual one, namely $Inv\Gamma(a, 1/b)$ instead of $Inv\Gamma(a, b)$ which made the calculation less stable and it took us some time to notice it.

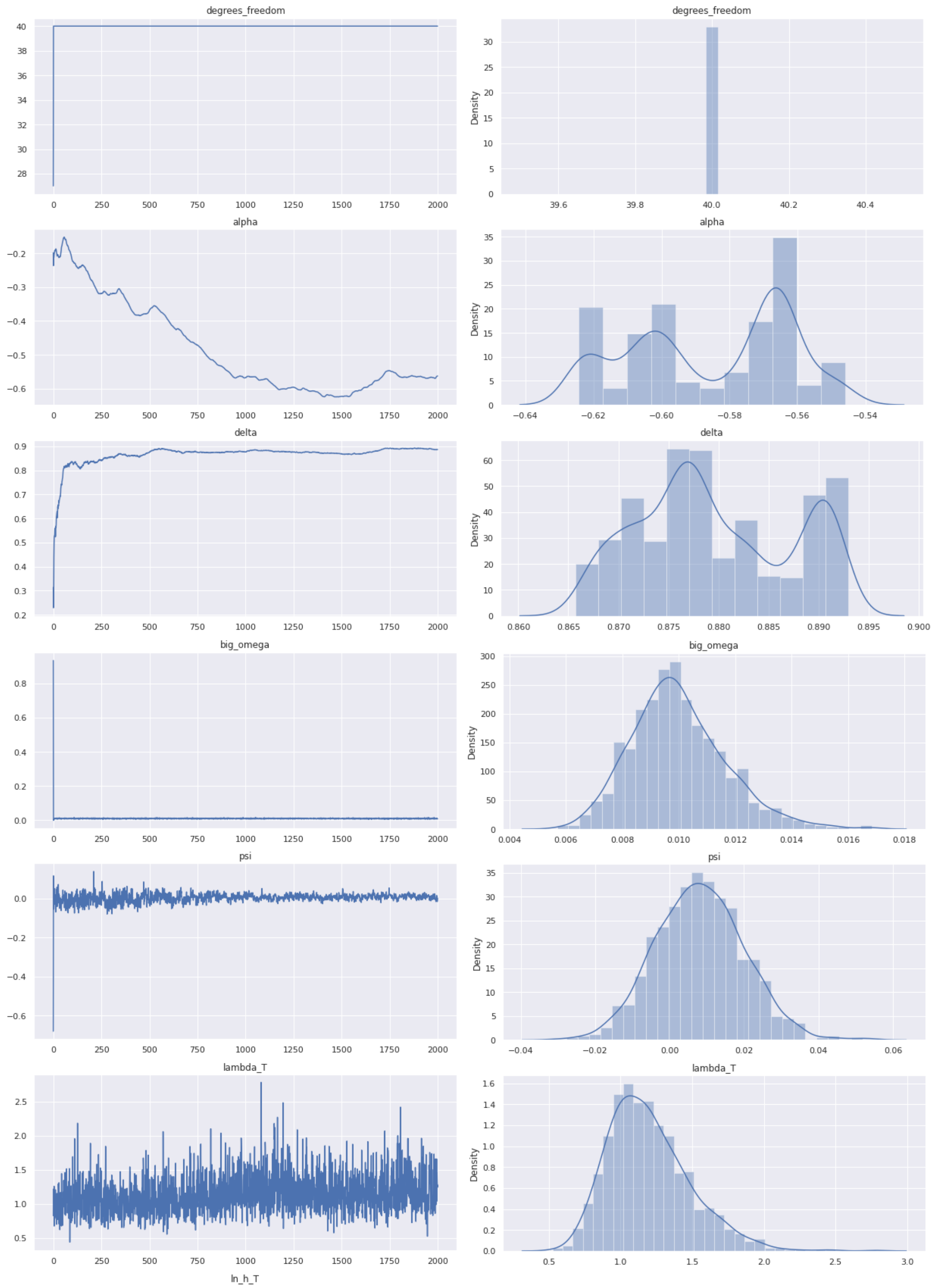

Figure 2: Trajectories and posteriors for the full model and simulated data

2.2 Real data

We will now test the basic and full model real data. Paradoxally, we obtain better results when applied to real data.

2.2.1 Basic model

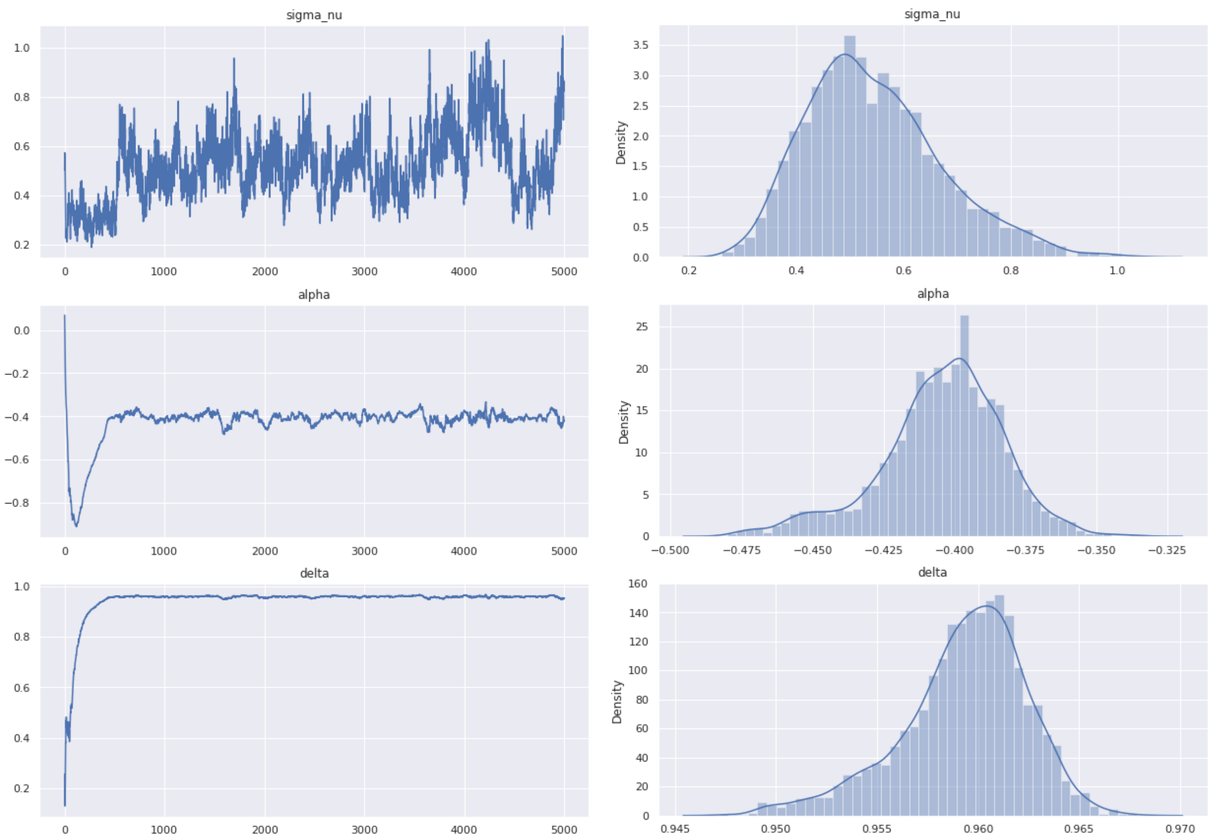

Figure 3: Trajectories and posteriors for the basic model and real data

We can see that α and δ behave nicely, as in the basic model and the problems with σ_ν have worsened : it doesn't seem to converge anymore

Looking at their distribution, α seems to be close to a normal distribution, σ_ν to an inverse gamma and δ is more skewed toward the left.

For parameter estimation, we obtain :

- σ_ν has mean 0.55 and std 0.13
- α has mean -0.40 and std 0.21
- δ has mean 0.96 and std 0.0031

2.2.2 Full model

Again, the look of the result is similar to the case with simulated data, but worse. Moreover, we can notice that ψ is not drifting but its variance is increasing over time.

- Ω has mean 0.011 and std 0.0033
- α has mean -0.49 and std 0.027
- δ has mean 0.92 and std 0.021
- ψ has mean 0.01 and std 0.011
- ν is 3

We can see that this new model was able to capture some stylised effects such as the fat-tailed distribution (as $\nu = 3$). The leverage effect isn't shown for the studied time period (as ψ remains low and positive).

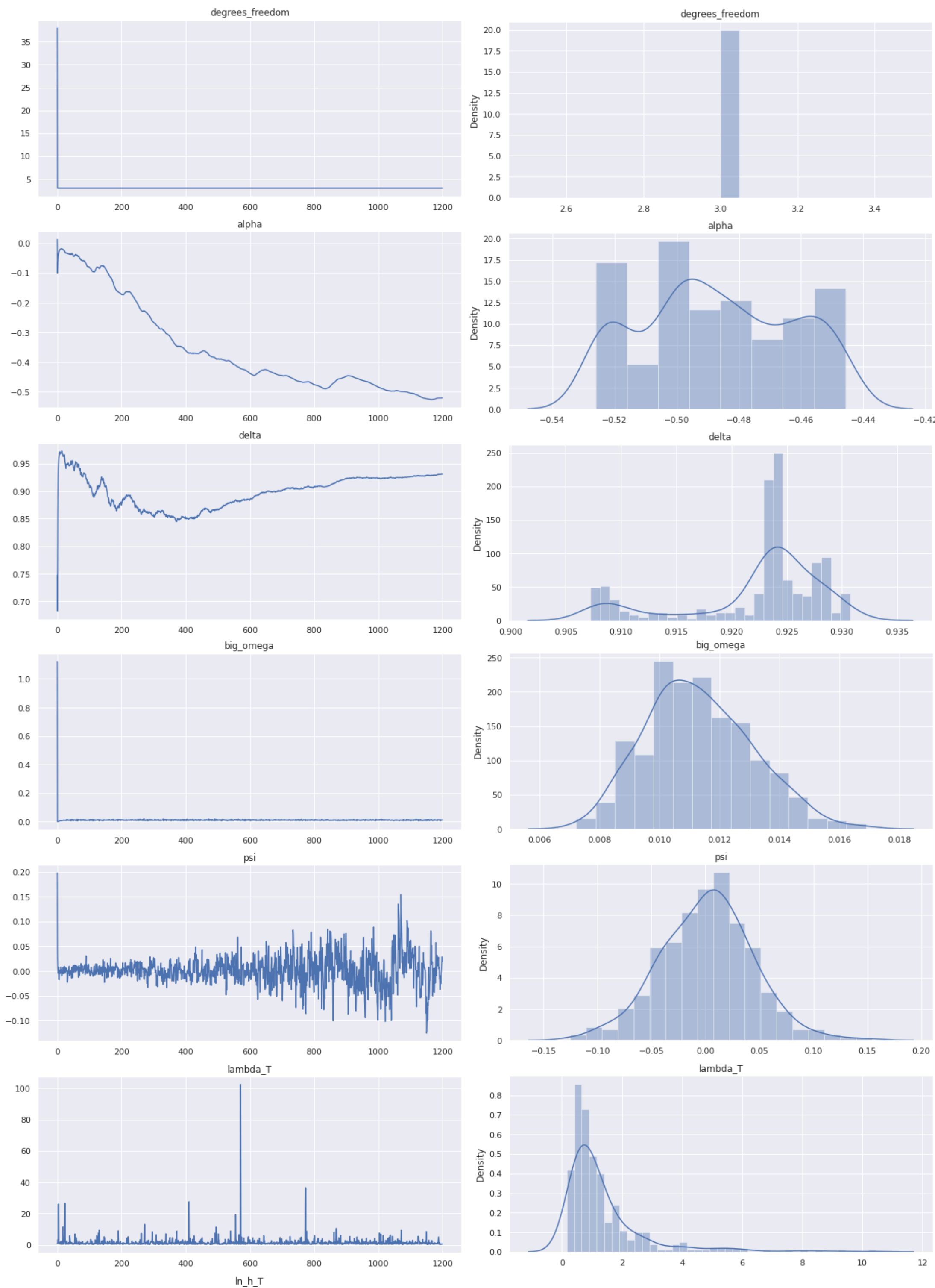

Figure 4: Trajectories and posteriors for the full model and real data